



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2000

Die Übertragbarkeit der Evaluationsstandards auf unterschiedliche Anwendungsfelder

Widmer, Thomas ; Beywl, Wolfgang

Abstract: Anschließend an die Veröffentlichung der übersetzten ‚Program Evaluation Standards‘ (nachfolgend kurz ‚Standards‘ genannt) des ‚Joint Committee an Standards for Educational Evaluation‘ (kurz ‚Joint Committee‘) wird die Fachwelt erörtern, ob, mit welchen Einschränkungen und welchen Ergänzungen die ‚Standards‘ auf konkrete Evaluationen in unterschiedlichen Politikbereichen im deutschsprachigen Raum anwendbar sind. Ohne den Ergebnissen dieser Diskussion vorgreifen zu wollen, möchten wir im folgenden Argumente anführen und Quellen benennen, die sich mit den Fragen der Anwendbarkeit und Übertragbarkeit der ‚Standards‘ beschäftigen und eine eigene Position formulieren.

DOI: https://doi.org/10.1007/978-3-322-93234-1_8

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-157848>

Book Section

Accepted Version

Originally published at:

Widmer, Thomas; Beywl, Wolfgang (2000). Die Übertragbarkeit der Evaluationsstandards auf unterschiedliche Anwendungsfelder. In: Sanders, James; Beywl, Wolfgang; Widmer, Thomas. Handbuch der Evaluationsstandards: die Standards des "Joint Committee on Standards for Educational Evaluation". Wiesbaden: VS Verlag für Sozialwissenschaften, 243-257.

DOI: https://doi.org/10.1007/978-3-322-93234-1_8

Die Übertragbarkeit der Evaluationsstandards auf unterschiedliche Anwendungsfelder

Thomas Widmer und Wolfgang Beywl

1. Einleitung und Fragestellung

Anschließend an die Veröffentlichung der übersetzten ‚Program Evaluation Standards‘ (nachfolgend kurz ‚Standards‘ genannt) des ‚Joint Committee on Standards for Educational Evaluation‘ (kurz ‚Joint Committee‘) wird die Fachwelt erörtern, ob, mit welchen Einschränkungen und welchen Ergänzungen die ‚Standards‘ auf konkrete Evaluationen in unterschiedlichen Politikbereichen im deutschsprachigen Raum anwendbar sind. Ohne den Ergebnissen dieser Diskussion vorgreifen zu wollen, möchten wir im folgenden Argumente anführen und Quellen benennen, die sich mit den Fragen der Anwendbarkeit und Übertragbarkeit der ‚Standards‘ beschäftigen und eine eigene Position formulieren.

Diese Thematik soll entlang folgender vier, in ihrer Abfolge zunehmend weiter greifenden Fragen erörtert werden, wobei wir den beiden letzten Fragen eigene Abschnitte widmen:

1. Wie lassen sich die ‚Standards‘ auf konkrete, oft einzigartige Evaluationen anwenden? (Anwendbarkeit auf spezifische Evaluationen)
2. Auf welche Gegenstandsarten (Programme, Projekte, Materialien usw.) sind die ‚Standards‘ gemäß ihrem Selbstverständnis anwendbar? (Anwendbarkeit auf verschiedene Gegenstandsarten)
3. Wie ist die Anwendbarkeit der ‚Standards‘ über ihren Entstehungskontext, den Bereich „Bildung und Erziehung“ hinaus auf andere Gegenstandsfelder zu sehen? (intersektorale Übertragbarkeit)
4. Welche Möglichkeiten und Grenzen gibt es bei der Übertragung aus der US-amerikanischen auf europäische und speziell deutschsprachige Gesellschaften? (internationale Übertragbarkeit)

Vorrangig geht es in diesem Beitrag (siehe Übersicht 1) um Anwendungskontext B (Diese Frage wird in Abschnitt 4 diskutiert) und um Anwendungskontext C (siehe Abschnitt 5). In Abschnitt 6 wenden wir uns der Frage einer generellen, also gleichzeitig internationalen wie intersektoralen Übertragbarkeit der ‚Standards‘ zu. Es wird erörtert, ob die ‚Standards‘ auch im Anwendungskontext D eingesetzt werden können. Dabei werden wir uns einerseits auf theoretische Überlegungen und andererseits auf empirische Evidenz stützen. Abschnitt 7 enthält schließlich eine Zusammenfassung der Resultate aus den vorangehenden Überlegungen.

Übersicht 1: Anwendungsfelder der ‚Standards‘

	<i>Bildung und Erziehung</i>	<i>Andere Gegenstandsfelder</i>
<i>USA</i>	Entstehungskontext A	Anwendungskontext B
<i>Andere Staaten (bes. D, A & CH)</i>	Anwendungskontext C	Anwendungskontext D

Wenn wir im folgenden von den ‚Standards‘ sprechen, meinen wir den kompletten Inhalt des Buches der amerikanischen Originalversion; in dieser Publikation ist dies der Teil A.

Kernbereich sind die 30 Standard-Kapitel. Diese bestehen jeweils aus sechs Teilen: der Standard-Bezeichnung, einer in einem Satz formulierten Soll-Aussage (zusammengefaßt von uns als „Standard im engeren Sinne“ bezeichnet), dem „Überblick“, den „Richtlinien“, den „Häufigen Fehlern“ und den „Anschauungsbeispielen“ (vgl. genauer oben, S. 25)

Zuerst werden wir anschließend kurz darauf eingehen, welche Reaktionen die ‚Standards‘ im allgemeinen hervorgerufen haben. Dabei wird also thematisiert, wie die ‚Standards‘ in der Fachwelt aufgenommen wurden, wobei der Schwerpunkt dieser Ausführungen auf der Akzeptanz im Entstehungskontext A, also dem US-amerikanischen Bildungs- und Erziehungsbereich, liegen wird.

2. Rezeption und Bewährung der ‚Standards‘ in ihrem Entstehungskontext

Voraussetzung für die ernsthafte Diskussion darüber, die ‚Standards‘ auf Anwendungsfelder *außerhalb* ihres Entstehungskontextes zu übertragen, ist, daß sich die ‚Standards‘ *in* ihrem Entstehungskontext bewährt haben. Dem vorausgesetzt ist eine hinreichend breite und positive Rezeption der ‚Standards‘ in der Fachwelt.

Die ‚Standards‘ sind seit ihrem Erscheinen in Evaluationsfachkreisen auf große Beachtung gestoßen. Sie werden in einschlägigen Lehrbüchern (Weiss 1998:110, Patton 1997:17, Posavac und Carey 1997:86, Worthen et al. 1997:441-5, Owen and Rogers 1999:151-69), Nachschlagewerken (Scriven 1991:155 und 230), Handbüchern (Greene 1994:535, Hatry et al. 1994: 592) und weiteren Hauptwerken der Evaluationsliteratur (House 1993:164-5, Guba und Lincoln 1989:229-33; vgl. auch die Beiträge in Stark and Thomas 1994) diskutiert. Sie werden genutzt zur Schwachstellenanalyse (Newman and Brown 1986), aber auch als Leitlinie zur Verbreitung von Evaluationsergebnissen (Torres et al. 1996:57, 66 & 158), als Bewertungskriterien im Rahmen von Meta-Evaluationen (Horvath 1995, Halpern 1988:39-40), zur Bewertung der Evaluationen der Jahrestagungen der American Evaluation Association (Foley and Newell 1995) oder zur Einschätzung neuerer Entwicklungen der Evaluationsmethodologie (Sanders 1997:402, Fetterman 1996:3

und 1995, Stufflebeam 1994). Worthen (1994:11) deutet sie als wichtigen Indikator der Professionalisierung der Evaluation. Verschiedene Buchbesprechungen weisen auf die Bedeutung der ‚Standards‘ hin (Hansen 1994, Patton 1994a&b). Wie diese – längst nicht vollständige – Übersicht zeigt, nimmt die Evaluationsliteratur die ‚Standards‘ intensiv zur Kenntnis und nutzt sie rege für Zwecke der Evaluationsausbildung, der Steuerung von Evaluationen und schließlich auch der Meta-Evaluation. Die Reaktionen sind großmehrheitlich positiv. Es gibt einige etwas kritischere Stimmen, die sich meist auf bestimmte Aspekte der ‚Standards‘ beziehen (Stake 1981, Hansen 1994: 560-1, Love 1994: 32-3, punktuell Patton 1994a). Die ‚Standards‘ finden nun auch im vielbeachteten Lehrbuch von Rossi, Freeman und Lipsey (1999: 425-6) Erwähnung. Die Autorengruppe mit dem inzwischen emeritierten Nestor der Programmevaluationsforschung, dem international angesehenen Soziologieprofessor Peter H. Rossi, sieht ein verbindliches und wirksames Set von Evaluationsstandards und ethischen Leitprinzipien noch in Entwicklung begriffen, auf der Basis der ‚Standards‘ einerseits, der Leitprinzipien der American Evaluation Association andererseits. Allerdings bedürfe es aus ihrer Sicht zusätzlich einer auf ‚Fallrechtsprechung‘ (*adjudicated case law*) beruhenden Schiedspraxis, die einen verbindlichen und sanktionsbewehrten Rahmen für gute Praxis von Evaluatoreninnen darstellen. Bis dieses Schieds- und Sanktionssystem aufgebaut sei, wofür sich gegenwärtig weder im Joint Committee noch bei der AEA Ansätze fänden, müßten Evaluatoreninnen sich auf die eher allgemeinen Grundsätze beziehen, wie sie in den ‚Leitlinien‘ und den ‚Standards‘ formuliert sind.

Die ‚Standards‘ fließen auch zunehmend bei öffentlichen Stellen in die Evaluationsgestaltung mit ein. So setzen die US-Bundesstaaten Louisiana, Hawaii und Florida die ‚Standards‘ als Leitfaden für Evaluationen in Bildung und Erziehung, aber auch im öffentlichen Sektor im allgemeinen ein (Worthen/Sanders/Fitzpatrick 1997:445).

Zur insgesamt positiven Aufnahme der ‚Standards‘ haben unseres Erachtens die folgenden Aspekte wesentlich beigetragen (vgl. dazu auch Widmer 1996:9-10):

- *Method(olog)ische Offenheit*: Die ‚Standards‘ sind nicht einer spezifischen methodologischen und methodischen Schule verpflichtet, sondern folgen einer offenen, integrativen Perspektive. So sind z.B. weder hypothesen-/theoriegeleitete noch naturalistische Ansätze, weder qualitative noch quantitative Methoden diskriminiert oder gar ausgeschlossen.
- *Umfassender Ansatz*: Die ‚Standards‘ sind darauf ausgerichtet, die Qualität einer Evaluation als Ganzes zu erfassen. Sie konzentrieren sich nicht auf spezifische Aspekte einer Evaluation, wie etwa die Qualifikationen oder das ethische Handeln der Evaluatoreninnen, die methodische Qualität von Datenerhebungsinstrumenten oder die Sicherung der Prozeßqualität einer Evaluation.
- *Breite institutionelle Abstützung*: Durch die aktive Mitwirkung einer Reihe bedeutender Fach- und Berufsorganisationen sowie durch den breit angelegten Erarbeitungsprozeß unter Einbezug einer Vielzahl von fachkundigen Personen konnte eine hohe Akzeptanz bei den Adressaten der ‚Standards‘ erreicht werden.

- *Hoher Detaillierungsgrad:* Die ‚Standards‘ verfügen mit insgesamt dreißig Einzelstandards, darunter jeweils mehreren „Richtlinien“ und aufgeführten „Häufigen Fehlern“, über eine hohe Differenziertheit, die es ermöglicht, Evaluationen in zahlreichen Facetten zu betrachten. Die ‚Standards‘ bieten damit detaillierte Kriterien, um fundierte Aussagen zur Qualität einer Evaluation zu machen. Sie können bei der Durchführung von Evaluationen für viele der erforderlichen Teilschritte handlungsleitend sein und bei der Lösung kritischer Probleme Orientierung bieten.
- *Praktische Verwendbarkeit:* Die ‚Standards‘ fallen zwar mit über 200 Druckseiten recht umfangreich aus. Die Zahl von dreißig Einzelstandards und der Umfang des jeweiligen Textes scheinen akzeptiert zu werden und die Praktikabilität der ‚Standards‘ nicht einzuschränken. Dies wird erleichtert durch die vorhandenen Find- und Verweissysteme (wie z.B. das funktionale Inhaltsverzeichnis).

Diese Charakteristika in der Erarbeitung und konzeptionellen Gestaltung treffen offensichtlich – zumindest in ihrem Entstehungsland – auf einen hohen Bedarf an Bewertungsmaßstäben für Evaluationen, die bei der Politikgestaltung und in vielen Politikfeldern an der Schwelle zum 21. Jahrhundert eine erhöhte Relevanz zugemessen bekommen:

„... evaluation – not only for the purposes of accountability and good management, but also for knowledge building and sharing, for institutional learning and development, for governmental and democratic reform through the serious examination of public policy – has become a precious and unique tool as we prepare to deal with the new socioeconomic, political and infrastructure needs of the next century.“ (Chelimsky 1997: 6).

An den vielfältigen positiven Reaktionen auf die ‚Standards‘ läßt sich erkennen, daß die US-amerikanische ‚Evaluationsgemeinde‘ die ‚Standards‘ als ein Instrument anerkennt, das einen maßgeblichen Beitrag zur Professionalisierung der Evaluation leisten kann.

3. Anwendbarkeit auf spezifische Evaluationen und auf verschiedene Gegenstandsarten

Wie sind die ‚Standards‘ auf den je spezifischen Fall einer Evaluation oder auf eine Gruppe von Evaluationsstudien (bspw. zu verschiedenen lokalen Programmen im Rahmen eines landesweiten Modellprogramms) abzustimmen? Im Kapitel ‚Anwendung der Standards‘ wird deutlich darauf hingewiesen, daß die ‚Standards‘ auf die Bedürfnisse angepaßt werden müssen, die im konkreten Einzelfall bestehen (vgl. oben S. 32). Das heißt, daß die ‚Standards‘ auch in ihrem Entstehungskontext – Bildung und Erziehung in den USA – nicht ohne Anpassungen auf die konkrete Evaluationsstudie eingesetzt werden können. Aus ihrem Selbstverständnis stellen die ‚Standards‘ somit einen verbindlichen Rahmen dar, innerhalb dessen notwendige Adaptionen an spezifische Evaluationsgegenstände,

kulturelle, regionale, organisatorische oder finanzielle Bedingungen konzeptionell vorgesehen sind.

Dies stellt eine gute interne Voraussetzung dafür dar, die ‚Standards‘ auch auf andere Gegenstandsfelder oder politische Systeme zu übertragen. Von einer Anwendung der ‚Standards‘ in anderen Kontexten müßte allerdings dann Abstand genommen werden, wenn spezifizierte Gründe bestünden, die gegen eine solche Übertragung sprechen. Es sei darauf hingewiesen, daß die in den ‚Standards‘ enthaltenen Anschauungsbeispiele (Fallbeispiele und -analysen) in anderen Kontexten darauf zu prüfen sind, ob die Bedingungen zu ihrer analogen Übertragung gegeben sind. Häufig beziehen sie sich auf Evaluationen im öffentlichen Schulsystem der USA, welches sich von dem in Deutschland, Österreich oder der Schweiz in vielerlei Hinsicht unterscheidet, insbesondere was seine Finanzierung und die Gestaltung der Schulaufsicht betrifft (vgl. Hadden 1992: 605)³³. Insbesondere, wenn gesetzliche Regelungen betroffen sind, ist die Übertragbarkeit der Fallbeispiele erheblich begrenzt.

Zunächst sei ein mögliches Mißverständnis angesprochen, das sich aus der Benennung „Standard“ ergeben kann. Owen und Rogers (1999:163) weisen darauf hin, daß die Bezeichnung „Standards“ in vielen Zusammenhängen mit einzuhaltenen (Minimal-)Kriterien verbunden ist. Solche (Minimal-)Standards müssen operational definiert sein, um entscheiden zu können, ob die Vorgabe eingehalten wurde oder nicht. Ein ganz ähnliches Verständnis von „Standards“ hat das Qualitätsmanagement: Ein Qualitätsstandard ist dort eine nicht zu unterschreitende Qualitätsforderung an die Beschaffenheit eines materiellen oder immateriellen Produktes (Dienstleistung). Die Deutsche Gesellschaft für Qualität weist in ihren Begriffsbestimmungen zum Qualitätsmanagement ausdrücklich darauf hin, daß diese Qualitätsforderungen verbindlich sind (DGQ 1995:26). Grillenberger (1996: 24) nennt in einem Owen und Rogers analogen Verständnis als Beispiel eines Qualitätsstandards: „95% aller zugestellten Pakete sollen pünktlich zugestellt sein.“

Keinesfalls stellen die ‚Standards für die Evaluation von Programmen‘ in diesem Sinne operationalisierte, rigide, in unterschiedlichen Kontexten einzuhalten-de Normen dar. Sie entsprechen eher dem in Fuchs u.a. (1975: 651) dargestellten Verständnis von „Standard“ als einem „Maßstab zur Bewertung eines Verhaltens als gut, wünschenswert o.ä.“, also einer sozialen, in Aushandlungsprozessen zu konkretisierenden Richtlinie. Ähnlich lautet die Definition aus dem Glossar der ‚Standards‘: „Ein von Experten allgemein anerkanntes Prinzip, das bei der Durchführung und Nutzung von Evaluationen als Maßstab genutzt wird, um den Wert oder die Qualität einer Evaluation einzuschätzen.“ (s. o. S. 232)

Die künftige Fachdiskussion wird zu klären haben, ob sie auch im deutschsprachigen Raum bei der – sicher mißverständlichen – Bezeichnung „Standards“ bleiben will oder andere Bezeichnungen vorzieht.³⁴

33 Die Unterschiede zwischen Deutschland und der Schweiz sind hier jedoch oft nicht minder gering; (s.u., Abschnitt 4).

34 Owen und Rogers (1999: 163) plädieren z.B. für den Begriff „*Codes of Behaviour*“, ins deutsche vielleicht als „Verhaltenskodizes“ zu übersetzen, wobei mit dieser Bezeichnung wieder-

Ein weiteres mögliches Mißverständnis birgt der Untertitel der ‚Standards‘: „How to Assess Evaluations of Educational Programs“. Hieß es in der ersten Auflage aus dem Jahre 1981 noch „Standards for Evaluations of Educational Programs, Projects and Materials“, so sind im neuen Untertitel lediglich noch ‚Programme‘ aufgeführt. Weist dies auf eine Einschränkung *des zu evaluierenden Gegenstandes* (nachfolgend auch Evaluandum genannt) auf die Gegenstandsart „Programm“, die Ausgrenzung z.B. von Maßnahmen, Projekten, Materialien oder ganzer Policies hin?

Zu dieser Frage sind unsererseits keine umfangreichen Ausführungen erforderlich, enthalten doch die ‚Standards‘ dazu klärende Hinweise. In der Einführung zu den ‚Standards‘ (vgl. oben, 23ff) weist das ‚Joint Committee‘ deutlich darauf hin, daß sich die ‚Standards‘ auf Evaluation von Programmen, Projekten und Materialien beziehen. Der Terminus „Programm“ wird also als übergreifende Bezeichnung genutzt, die auch andere durch ein Konzept gesteuerte Bündel von Interventionen, Prozessen, Ergebnissen einschließt.³⁵ Explizit ausgeschlossen werden lediglich Evaluationen, die sich mit Personen befassen. Für diesen Typ von Evaluationen hält das ‚Joint Committee‘ eine andere Publikation bereit, nämlich ‚The Personnel Evaluation Standards‘, die im Jahre 1988 vom ‚Joint Committee‘ herausgegeben wurden (Joint Committee 1988).

4. Intersektorale Übertragbarkeit der ‚Standards‘

Die Frage, der wir uns in diesem Abschnitt detaillierter zuwenden, betrifft den *sachlichen Geltungsbereich* der ‚Standards‘. Das ‚Joint Committee‘, das sich – wie dies in der Bezeichnung ‚The Joint Committee on Standards for Educational Evaluation‘ explizit zum Ausdruck kommt – mit ‚Standards‘ für Evaluationen im Bereich Bildung und Erziehung befaßt, hat die ‚Standards‘ entwickelt. Während sie in der ersten Auflage aus dem Jahre 1981 noch den Titel ‚Standards for Evaluations of Educational Programs, Projects, and Materials‘ trugen, wurde die 1994 erschienene zweite Ausgabe mit dem Haupttitel ‚The Program Evaluation Standards‘ überschrieben. Die Einschränkung auf Bildung und Erziehung ist in der zweiten Ausgabe nur noch in dem Untertitel ‚How to Assess Evaluations of Educational Programs‘ enthalten. Es stellt sich nun die Frage, inwiefern die ‚Standards‘ auch auf Evaluationen außerhalb des Erziehungs- und Bildungsbereichs angewandt werden können.

Betrachtet man die ‚Standards‘ als Ganzes, ist – wie bereits erwähnt – zwischen den Abschnitten Standard, Überblick, Richtlinien und ‚Häufige Fehler‘ einerseits

um die Gefahr einer akteurorientierten Deutung besteht (vgl. dazu den Beitrag „Die ‚Standards‘ im Vergleich mit anderen Regelwerken zur Qualität fachlicher Leistungserstellung“ in diesem Band).

35 Verwiesen sei auf die häufig zitierte Definition von Smith (1989) für ein Programm: „Set of planned activities directed toward bringing about specified change(s) in an identified and identifiable audience.“

und den illustrierenden Fallbeschreibungen und -analysen andererseits zu unterscheiden. Letztere sind durchwegs im Bildungs- und Erziehungsbereich in einem weiteren Sinne lokalisiert und sind auch bezüglich der genutzten Sprache und den fachlichen Perspektiven diesem Kontext verhaftet.

Die Übersicht 2 listet differenziert auf, aus welchen Gegenstandsfeldern die insgesamt 53 Anschauungsbeispiele aus der ‚Standards‘-Publikation stammen: Dominant ist die Schule (Elementar- und Sekundarschule), gefolgt von der Hochschule, der betrieblichen Bildung und der sozialen Arbeit.

Übersicht 2: Gegenstandsbereiche in den Anschauungsbeispielen der ‚Standards‘*

Gegenstandsbereich	Anschauungsbeispiele der ‚Standards‘
Elementarschulen	N1.1, N4.1, N7.1, D1.1, D2.1, D3.1, K3.1, K7.1, G6.1, G8.1 und G11.1
Sekundarschulen	N3, N5.1, N6.1, K1, K2.1, K5.1, K6.1, G1.1, G2.1, G3, G5.1, G7.1, G8.1 und G10
Colleges/Universitäten	N1.2, N2.1, N2.2, N6.2, D3.2, K3.2, K4.1, K5.2, K6.2, K7.2, K8, G4, G6.2, G7.1, G9.2 und G11.2
Betriebliche Bildung	N5.2, N7.2, K2.2, K4.2, G1.2, G2.2 und G12
Soziale Arbeit	N4.2, D2.2, G5.2 und G9.1

* Die Ziffer hinter dem Punkt bezeichnet jeweils die Nummer des Anschauungsbeispiels innerhalb eines ‚Standard‘-Kapitels im Teil A dieses Buches.

Auch wenn sich immerhin vier Anschauungsbeispiele mit sozialer Arbeit und nicht mit Bildung und Erziehung im engeren Sinne befassen, überwiegt der Entstehungskontext Schule/Hochschule doch eindeutig. In bezug auf die Anschauungsbeispiele kann also höchstens von einer Übertragbarkeit gesprochen werden, wenn sich die beispielhaft dargestellte Grundproblematik des jeweiligen Falles in anderen Kontexten nachweislich analog stellt.

Die Frage der Übertragbarkeit auf andere Politik- oder Gegenstandsfelder sehen wir hingegen bei den anderen fünf Teilen der Standard-Kapitel als offen an. Dies gilt besonders für die „Standards im engeren Sinne“, jeweils bestehend aus einer Standard-Bezeichnung und einer kurzen Soll-Aussage zur Ausgestaltung von Evaluationen. Inwiefern setzen sie Annahmen und Bedingungen voraus, die bildungs- oder erziehungsspezifisch sind?

Wie eine Durchsicht der Kurzfassung der ‚Standards‘, welche aus den insgesamt 30 „Standards in engeren Sinne“ besteht, ergibt, werden hier keine bildungs- oder erziehungsspezifischen Begriffe benutzt. Weder die Bezeichnungen von Personen noch solche von Strukturen, Prozessen oder Produkten lassen auf die Bereiche Bildung oder Erziehung schließen. „Evaluatorinnen“ oder „Beteiligte & Betroffene“ gibt es auch in allen anderen Feldern, in denen Evaluation stattfindet. Es wird nicht von „Schülern“ oder „Studierenden“ sondern von „Zielgruppen“, nicht vom „Lehrkörper“ sondern von den „Mitarbeitenden“ gesprochen. Dieser Wortgebrauch

läßt die ‚Standards‘ offen für Anwendungen über den Bereich Bildung und Erziehung hinaus.

Eine Ausnahme bildet die Richtlinie A) des Standards K1 „Unterstützung der Dienstleistungsorientierung“:

„A. Es sollten Evaluationen geplant werden, welche die Qualität von Programmen zur Erziehung, Aus- und Weiterbildung fördern.“

An dieser Stelle wird deutlich, daß dieser ‚Standard‘ dem Gegenstandsfeld Bildung und Erziehung entstammt. Eine Übertragung der in dieser Richtlinie enthaltenen Aussage fällt jedoch nicht allzu schwer. Die Richtlinie hat ja zum Zweck, die Evaluation nach den Zielsetzungen des Evaluationsgegenstandes (bzw. des Programms) auszurichten. So soll die Evaluation einen Beitrag zur Förderung der Qualität des Evaluationsgegenstandes leisten. Dieser Grundsatz läßt sich genauso auf Programme außerhalb von Bildung und Erziehung anwenden, insbesondere bei anderen personenbezogenen Dienstleistungen.

Übersicht 3: Im ‚Joint Committee‘ vertretene Organisationen, 1981 und 1994

1981	1994
American Association of School Administrators	American Association of School Administrators
American Educational Research Association	American Educational Research Association
	American Evaluation Association
American Federation of Teachers	American Federation of Teachers
American Personnel and Guidance Association	
American Psychological Association	American Psychological Association
	Association for Assessment and Counseling
Assoc. for Supervision and Curriculum Development	Assoc. for Supervision and Curriculum Development
	Canadian Society for the Study of Education
Council for American Private Education	
Education Commission of the States	
	Council of Chief State School Officers
	Council on Postsecondary Accreditation
National Association of Elementary School Principals	National Association of Elementary School Principals
	National Association of Secondary School Principals
National Council on Measurement in Education	National Council on Measurement in Education
National Education Association	National Education Association
National School Boards Association	National School Boards Association

Quelle: ‚Joint Committee‘ 1981: viii und 1994: xi-xii

Betrachtet man die institutionelle Abstützung der ‚Standards‘, ist zunächst die Frage von Interesse, welche Organisationen sich an der Ausarbeitung der ‚Standards‘ beteiligt haben. Die Übersicht 3 listet die Organisationen auf, welche im ‚Joint Committee on Standards for Educational Evaluation‘ bei der Ausarbeitung der ersten und zweiten Fassung der ‚Standards‘ (1981 resp. 1994 veröffentlicht) vertreten waren.

Wie die Übersicht zeigt, sind viele der beteiligten Organisationen im Bildungs- und Erziehungsbereich verankert. Es sind aber auch einige Organisationen vertreten, die nicht ausschließlich in Bildung und Erziehung tätig sind. In erster Linie sind dies die ‚American Psychological Association‘ (APA) und, bei der Erarbeitung der zweiten Auflage, die ‚American Evaluation Association‘ (AEA). Bei der APA handelt es sich um eine etablierte, disziplinär orientierte Wissenschaftsvereinigung. Die AEA stellt mit 2200 Mitgliedern die größte Evaluationsgesellschaft in den USA (und damit weltweit) dar, deren Mitglieder zwar häufig, aber bei weitem nicht immer im Bereich Bildung und Erziehung beschäftigt oder dort als externe Evaluatoren tätig sind. In einem ihrer Publikationsorgane ‚The American Journal of Evaluation‘ (vormals ‚Evaluation Practice‘) wird eine Kolumne zur Diskussion ethischer Herausforderungen geführt. Darin wird explizit großen Wert darauf gelegt, daß auf die ‚AEA Guiding Principles‘ (siehe Shadish et al. 1995 und weiter unten, S. 282f) und auf die ‚Standards‘ Bezug genommen wird (siehe Morris 1998).

Vergleicht man schließlich die Fassung von 1981 mit der von 1994, fällt hinsichtlich der institutionellen Abstützung auf, daß die Ausgabe aus dem Jahre 1994 durch das American National Standards Institute (ANSI, siehe <http://www ansi.org>) als ‚American National Standard‘ anerkannt wurde. Zwar bezieht sich diese Anerkennung explizit auf den zentralen Geltungsbereich der Standards, also Bildung und Erziehung. Da die ‚Standards‘ aber die einzigen Evaluationsnormen im sozialwissenschaftlichen Bereich sind, die durch das ANSI anerkannt wurden, entfaltet dies Wirkungen über den Bildungsbereich hinaus.

Eine Studie, welche der Frage nachging, inwiefern die ‚Standards‘ für Evaluationen in der US-amerikanischen Geschäftswelt angemessen sind, kommt zu eindeutigen Ergebnissen. Die beigezogenen Expertenmeinungen aus Bildung/Erziehung sowie Geschäftswelt weisen in ihren Einschätzungen der ‚Standards‘ keine Differenzen auf (Stockdill 1986). Eine weitere Untersuchung, die sich mit der Entwicklung von Qualitätskriterien für Evaluationen in den unterschiedlichsten Gegenstandsfeldern befaßt, zieht die ‚Standards‘ bei, ohne auf Probleme bei der Anwendung auf Bereiche außerhalb von Bildung und Erziehung zu stoßen (Caracelli and Riggins 1994).

In diesen Anwendungen, die nicht speziell im Bereich Bildung und Erziehung sondern in anderen Gegenstandsfeldern erfolgen, sind Hinweise darauf ausgeblieben, daß die ‚Standards‘ nicht einsetzbar wären. Zwar waren in allen diesen Fällen Anpassungen der ‚Standards‘ erforderlich. Da dies jedoch auch für Evaluationen im Entstehungskontext nötig ist (s.o.), bestehen keinerlei Hinweise, die grundsätzlich gegen eine Anwendung der ‚Standards‘ außerhalb des Bildungs- und Erziehungsbereichs sprechen.

Aufgrund der uns vorliegenden Erkenntnisse theoretischer und empirischer Art kann die Übertragbarkeit der ‚Standards‘ von dem Entstehungskontext Bildung und Erziehung auf andere thematische Anwendungsfelder somit als grundsätzlich gegeben eingeschätzt werden.

5. Internationale Übertragbarkeit der ‚Standards‘

Hier soll die Frage erörtert werden, inwiefern die im US-amerikanischen Kontext entwickelten ‚Standards‘ in anderen Regionen, wie beispielsweise in Deutschland, Österreich oder der Schweiz, anwendbar sind. Es geht also um die *internationale Übertragbarkeit* der ‚Standards‘. Diese Thematik ist schon bei der rein sprachlichen Übertragung der ‚Standards‘ vom Amerikanischen ins Deutsche allgegenwärtig – für viele Termini der amerikanischen Evaluationsfachsprache gibt es keine unmittelbaren Entsprechungen im Deutschen.³⁶ Zur Klärung dieser Frage ist aber auch zu prüfen, ob sich aufgrund unterschiedlicher Kontextbedingungen Anpassungen der ‚Standards‘ aufdrängen oder gar generell von einer Anwendung der ‚Standards‘ in den deutschsprachigen Ländern abzusehen ist.

In den allgemeinen Teilen der 30 Standard-Kapitel (also den jeweils ersten fünf Teilen ohne die Anschauungsbeispiele) ist uns in der Gruppe der Korrektheitsstandards eine Aussage aufgefallen, die auf eine US-amerikanische Besonderheit verweist: Der Standard K6 – Offenlegung der Ergebnisse – thematisiert das „Recht auf Informationen („right-to-know“), ein mit dem Freedom of Information Act zusammenhängendes Recht z.B. auf Akteneinsicht bei Behörden“³⁷. Hier ist somit eine leichte Einschränkung gegeben.

Wie aus der Übersicht 3 (siehe oben) ersichtlich, handelt es sich bei den Organisationen, die bei der Erarbeitung der ‚Standards‘ im ‚Joint Committee‘ vertreten waren, nahezu ausschließlich um in den USA ansässige Organisationen. Eine Ausnahme bildet dabei im Jahre 1994 die ‚Canadian Society for the Study of Education‘. Neben einigen Personen aus internationalen Organisationen nahmen im Rahmen eines „International Validation Panel“ (vgl. Joint Committee 1994:197) unter anderem Experten aus Australien, Hongkong, Ägypten, Israel und Spanien an den Überarbeitungen für die Neuauflage teil. Offensichtlich liegt ein Bemühen des Joint Committee vor, die internationale Übertragung der ‚Standards‘ wenn nicht zu sichern, so doch zu erleichtern.

Außerhalb der USA wird in letzter Zeit auch in anderen Weltregionen vermehrt auf die ‚Standards‘ des ‚Joint Committee‘ verwiesen:

- So hat die Australasian Evaluation Society (AES) die ‚Standards‘ im August 1996 gutgeheißen. Die AES bezieht sich auch im Zusammenhang mit ihren Ethik-Richtlinien auf die ‚Standards‘ (<http://www.parklane.com.au/aes/ethics.htm>).

36 Z.B. stakeholder, formative, summative, merit, worth, outcome, impact.

37 Vgl. dazu die Ausführungen von Weiss (1990:100) und die Hinweise unter <http://www.aclu.org/library/foia.html#introduction> und <http://www.spj.org/foia/index.htm>

- Das Handbuch der Europäischen Union zu Qualitätsbewertungen von Evaluationsberichten zur Strukturpolitik enthält ebenfalls einen Hinweis auf die ‚Standards‘ (European Union 1996).

In der Literatur finden sich einige Hinweise zur Übertragbarkeit der ‚Standards‘ auf Evaluationen in anderen Ländern (vgl. dazu auch Hendricks and Conner 1995): Für Schweden wird die Anwendbarkeit der ‚Standards‘ nicht in Frage gestellt, auch wenn auf einige spezifische Aspekte hingewiesen wird, die für leichtere Anpassungen sprechen (Marklund 1985:356). Auch in Israel scheinen sich die ‚Standards‘ grundsätzlich anwenden zu lassen. Die Autoren weisen aber ebenfalls darauf hin, daß sie gewisse (übrigens nicht identische) Revisionen für erforderlich halten (Lewy 1984, Nevo 1985). Weitergehende Überlegungen zur Übertragbarkeit der ‚Standards‘ auf andere Gesellschaften enthält ein Artikel, der sich mit der Anwendbarkeit der ‚Standards‘ in Indien und Malta befaßt (Smith et al. 1993). Auch wenn dort weitergehende Anpassungen empfohlen werden, wird die Anwendung der ‚Standards‘ nicht grundsätzlich abgelehnt. Die Vorbehalte sind aber deutlich stärker als für andere untersuchte Länder (Smith et al. 1993:9-11). Die Autoren führen dies auf drei Faktoren zurück: Erstens zeichnen sich die Kulturen in den beiden Länder Indien und Malta dadurch aus, daß sie deutlich kollektivistischer orientiert sind, als die durch individualistische Werte geprägte US-Kultur. Zweitens sehen sie einen Unterschied darin, daß die Kulturen der beiden Länder eher assoziativ geprägt sind, im Gegensatz zur stärker abstrahierenden Kultur der USA (Smith et al. 1993:12). Ein dritter Punkt bezieht sich auf die Rolle in der fachlichen Urteilsbildung, welche Standards in Malta oder Indien zukommt – also in Kulturen, die ein deutlich von den USA zu unterscheidendes Bild von Professionalität aufweisen (Smith et al. 1993:7). Im Gegensatz zu den USA, wo Expertise als Ergebnis transparenter spezialisierter Verfahrensweisen den Kernbestandteil von Professionalität darstellt (vgl. dazu auch Worthen 1994), spielt diese in den Vergleichsländern Indien und Malta eine viel geringere Rolle. Dort steht die Glaubwürdigkeit der Person klar an erster Stelle. Sie vermag auch schwerwiegende Mängel im Bereich des systematischen, datenbasierten Nachweises aufzuwiegen.

Vergleichen wir nun die kulturellen Prämissen in den USA mit jenen in den deutschsprachigen Ländern Deutschland, Österreich und Schweiz im Hinblick auf die drei genannten Dimensionen kollektivistisch-individualistisch, assoziativ-abstrahierend und Professionalitätsauffassung, lassen sich keine wesentlichen Differenzen feststellen. Damit sprechen die Gegebenheiten in den drei genannten Dimensionen nicht gegen eine Anwendung der ‚Standards‘ im deutschsprachigen Raum, im Gegensatz zu anderen, nicht westlich geprägten Kulturräumen, wo Vorbehalte bestehen können.

Es finden sich einige praktische Anwendungen der ‚Standards‘ außerhalb der USA³⁸. So ist mit den ‚Standards‘ bei einer Qualitätsbewertung von Evaluationen in der schwedischen Entwicklungszusammenarbeit gearbeitet worden (Forss and Carlsson 1997).

38 Für Anwendungen der ‚Standards‘ im deutschsprachigen Raum, siehe auch die Hinweise im Abschnitt 6, S. 255f.

In Vergleichen zwischen der Evaluation in den USA und jener in Europa und im speziellen in den deutschsprachigen Staaten wird in der Literatur auf folgende Unterschiede hingewiesen:³⁹

- Größenverhältnisse,
- Politisches System (Föderalismus, direkte Demokratie, Rechtstradition und Stellung der Justiz, experimentelle Gesetzgebung),
- Entwicklungsstand der Sozialwissenschaften,
- Evaluationskultur.

Diese Aspekte verweisen sicherlich zunächst auf Unterschiede, die weniger auf die *anzustrebende* Qualität von Evaluationen ausgerichtet sind, sondern vielmehr auf die *bestehende* Qualität (und Quantität). Daß in den deutschsprachigen Ländern weniger evaluiert wird, als dies in den USA der Fall ist, braucht an dieser Stelle nicht betont zu werden. Daß die Häufigkeit von Evaluationen (etwa pro Milliarde Dollar Ausgaben der öffentlichen Haushalte) einen Einfluß auf die erreichte Evaluationsqualität haben kann, wird ebenfalls nicht bestritten – auch wenn sich dieser Zusammenhang sicherlich nicht in jedem Einzelfall wird erkennen lassen. Aber die hier zu beantwortende Frage ist eine andere: Führen die genannten Faktoren dazu, daß sich die Ansprüche an eine qualitativ hochstehende Evaluation grundsätzlich anders gestalten? Sind die Kriterien für die Qualität einer Evaluation – und damit auch die Gütekriterien – aus den angeführten Gründen grundlegend anders zu definieren? Oder anders ausgedrückt: Hier zu diskutieren gilt es nicht die Frage, ob die oben genannten Faktoren dazu führen, daß Evaluationen in den verschiedenen Ländern auf den identischen Qualitätsdimensionen andere Ausprägungen erreichen (können), sondern es geht darum, zu klären, ob andere Bewertungsdimensionen zugrunde gelegt werden sollten.

In einem Vergleich der Bildungspolitiken in Deutschland, Japan und der Schweiz, bei dem teilweise auch die USA als Vergleichsfall beigezogen werden, kommt Heidenheimer (1997:315-6) unter anderem zum Schluß, daß sich die Schulsysteme Deutschlands und der Schweiz in gewissen Aspekten deutlich voneinander unterscheiden, während das schweizerische durchaus mit jenem der USA vergleichbar ist und Japan eine Mittelstellung einnimmt. Dieses auf Anhieb überraschende Resultat zeigt, daß die Vermutung, es bestünden zwischen den Schulsystemen im deutschsprachigen Raum und jenen in den USA grundlegende Differenzen, bei einer differenzierteren Betrachtungsweise relativiert werden muß.

39 Siehe dazu ausführlicher Derlien 1997, Bussmann 1995, 1996 und 1997, sowie die Beiträge in Wagner et al. 1991, Horber 1990 und Rist 1990. Die in diesen Texten angeführten Auffassungen zu den Auswirkungen bspw. des politischen Systems auf die Evaluation teilen wir nicht in allen Fällen. Im Gegensatz zur dort postulierten Auffassung sind wir der Ansicht, daß der Evaluation in der direkten Demokratie die Aufgabe zukommt, zur Vorbereitung eines Plebiszites Informationen bereitzustellen. Weiter verfügen Abstimmungsergebnisse, die gemäß der in der zitierten Literatur vorgebrachten Auffassung auch die Funktion der Evaluation abdecken können, zumeist nicht über die zur zukünftigen Politikgestaltung erforderliche Informationsdichte.

In den Bereichen Bildung und Erziehung sind in Deutschland die ‚Standards‘ bislang nicht in nennenswertem Umfang eingesetzt worden. Vereinzelt wird auf sie Bezug genommen, z.B. bei Allgäuer (1997:41) auf die Definition von Evaluation. Die beiden an deutschen Hochschulen entstandenen Lehrbücher, die „Evaluation“ im Titel tragen, beschäftigen sich in einem Fall gar nicht mit den ‚Standards‘ (Bortz und Döring 1995); Wottawa und Thierau (1998:165-8) führen sie zwar im Literaturverzeichnis auf, ziehen sie jedoch im Kapitel „Bewertung sozialwissenschaftlicher Evaluation“ nicht bei. Auch Grohmann (1997), der eine umfassende Darstellung des „Problems der Evaluation in der Sozialpädagogik“ vornimmt und dabei auch zahlreiche amerikanische Quellen nutzt, übersieht die ‚Standards‘ als möglichen Bezugspunkt einer kriteriengeleiteten Bewertung von Evaluation.

In der Schweiz wäre etwa auf die Nutzung der ‚Standards‘ im ‚Programm zur Internen Qualitätsevaluation‘ (PIQ) hinzuweisen (Stamm 1998, bes. 42-45). Zwar wurden auch hier einige wenige Anpassungen vorgenommen, generelle Einwände gegen eine Anwendung der ‚Standards‘ auf Evaluationen von schweizerischen Schulen bestehen aber keineswegs.

Zusammenfassend läßt sich somit festhalten, daß sich die ‚Standards‘ in kulturellen Umfeldern, die mit jenem der USA vergleichbar sind, durchwegs als anwendbar erwiesen haben. Dies gilt insbesondere auch für die hier im Zentrum stehenden Länder Deutschland, Österreich und Schweiz. Etwas anders verhält es sich in Regionen, die eine größere kulturelle Distanz zu den USA aufweisen. In diesen Fällen kann nicht von vornherein von einer generellen Anwendbarkeit der ‚Standards‘ ausgegangen werden.

6. Anwendung der ‚Standards‘ außerhalb von Bildung und Erziehung im deutschsprachigen Raum

Wir haben die ‚Standards‘ in eigenen Forschungen⁴⁰ dahingehend überprüft, inwiefern eine empirische Anwendbarkeit in Feldern außerhalb des Bildungs- und Erziehungsbereichs in der Schweiz gegeben ist. Dabei wurden Evaluationen aus den unterschiedlichsten Themenbereichen berücksichtigt. Die Studien stammten aus Bereichen wie Umwelt-, Verkehrs- und Energiepolitik, Wirtschaftsförderung, Sozialpolitik, öffentliches Beschaffungswesen, Kulturpolitik, Agrarpolitik, europäische Integration. Mit dieser breiten Themenstreuung sollten die ‚Standards‘ auf ihre gegenstandsfeldunabhängige Anwendbarkeit überprüft werden. Die praktische Umsetzung der ‚Standards‘ erfolgte in der Form retrospektiver Meta-Evaluationen. Die beigezogenen Evaluationsstudien – insgesamt 15 – wurden dabei in detaillierten

40 Dabei handelt es sich um zwei Forschungsprojekte, die durch den Schweizerischen Nationalfonds (SNF) im Rahmen des Nationalen Forschungsprogramms ‚Wirksamkeit staatlicher Massnahmen‘ (NFP 27) finanziell unterstützt wurden (siehe Widmer 1996a-h sowie Widmer et al. 1996).

Fallstudien einer Qualitätsüberprüfung unterzogen. Die hier relevanten Ergebnisse⁴¹ aus diesen Studien sollen nun kurz zusammengefaßt werden: In keinem der fünfzehn Fälle stießen wir auf Hinweise, die gegen eine Anwendbarkeit außerhalb des Bildungs- resp. Erziehungssektors sprechen würden. Die ‚Standards‘ eigneten sich durchweg zur Beurteilung der Qualität der untersuchten Evaluationsstudien. Die für die analysierten Evaluationen zuständigen Evaluatorinnen waren nicht in allen Fällen mit der vorgenommenen Beurteilung einverstanden, wobei sich die Kritik⁴² jedoch in keinem der Fälle darauf bezog, daß ein sachfremdes Kriterienraster für die Meta-Evaluation beigezogen worden sei. Die Untersuchungen haben auch keinerlei Hinweise erbracht, daß sich die ‚Standards‘ nicht auf den schweizerischen Kontext übertragen ließen.

Aufgrund der Ergebnisse der genannten Forschungsprojekte sind bereits einige Anwendungen der ‚Standards‘ in der Schweiz zu beobachten. So verwendet die Eidgenössische Finanzkontrolle (EFK) eine leicht an die spezifische Situation adaptierte Fassung der ‚Standards‘ zur internen Qualitätsüberprüfung ihrer Wirtschaftlichkeitsprüfungen. Diese Qualitätsüberprüfung hat zum Ziel, die Qualität der durch die EFK erarbeiteten Evaluationen weiter zu verbessern. Zudem soll sie auf bestehende Schwachstellen hinweisen, um eine zielgerichtete interne Weiterbildung realisieren zu können. Die Wirtschaftlichkeitsüberprüfungen der EFK befassen sich mit allen Bereichen der Staatstätigkeit auf Bundesebene. Ein weiteres Kontrollorgan des Bundes, die Verwaltungskontrolle des Bundesrates (VKB), zieht die ‚Standards‘ in einer, ihren Bedürfnissen angepaßten Form ebenfalls für ihre Arbeit bei. Das schweizerische Bundesamt für Gesundheit (BAG) stützt sich in seinem ‚Leitfaden für die Planung von Projekt- und Programmevaluationen‘ auch auf die ‚Standards‘ (BAG 1997; vgl. dazu auch den Beitrag „Die ‚Standards‘ im Vergleich zu weiteren Regelwerken zur Qualität fachlicher Leistungserstellung“ in diesem Band). Eine vor kurzem in der Schweiz erschienene Einführung in die Politikevaluation nimmt ebenfalls Bezug auf die ‚Standards‘ (Klöti 1997:55-7, Widmer und Binder 1997:253-4). Weiter hat sich eine Arbeitsgruppe der Schweizerischen Evaluationsgesellschaft (SEVAL) mit der Erarbeitung von schweizerischen Evaluationsstandards befaßt und diese der SEVAL zur Beschlußfassung vorgeschlagen. Die Arbeitsgruppe, die sich aus Vertretern des Bundes, der Kantone, der Evaluationspraxis und der Wissenschaft zusammensetzt, hat als Grundlage für die Entwicklung der schweizerischen Evaluationsstandards ebenfalls die ‚Standards‘ beigezogen (Widmer/Landert/Bachmann 1999). Ein mit ähnlicher Zielsetzung 1998 gegründeter Arbeitskreis der Deutschen Gesellschaft für Evaluation (DeGEval) prüft ebenfalls die Nutzung der ‚Standards‘ für die Entwicklung eines Regelwerkes für die Evaluation (Beywl 1999).

41 Für weitere Untersuchungsergebnisse verweisen wir auf die ausführlichen Publikationen zu den beiden Forschungsprojekten (Widmer 1996a-h und Widmer et al. 1996).

42 Die für die evaluierten Evaluationen zuständigen Evaluatorinnen erhielten in den beiden Forschungsprojekten Gelegenheit, zu den in den Fallstudien enthaltenen Bewertungen Stellung zu nehmen. Diese Stellungnahmen sind als integraler Bestandteil der Fallstudien publiziert worden (siehe Widmer 1996a-h und Widmer et al. 1996).

7. Zusammenfassung

Aufgrund der dargestellten Überlegungen kommen wir zum Schluß, daß sich die ‚Standards‘ sowohl auf Evaluationen in Kontexten außerhalb des Bildungsbereichs wie auch auf solche außerhalb der USA anwenden lassen. Dabei ist die internationale Übertragbarkeit der ‚Standards‘ nur mit Einschränkungen gegeben. Während in demokratischen Industriestaaten westlicher Prägung mit ausreichenden rechtsstaatlichen Garantien einer Anwendung der ‚Standards‘ nichts entgegensteht, trifft dies für andere Staaten und Weltregionen nur bedingt zu.

Daß die ‚Standards‘ nicht unbesehen ohne Anpassungen an die konkrete Situation anwendbar sind, darüber besteht Einigkeit. Dies gilt nicht nur für den Einsatz der ‚Standards‘ in ‚fremden‘ Anwendungskontexten, sondern – wie dies das ‚Joint Committee‘ selbst festgehalten hat – auch für den Entstehungskontext der ‚Standards‘, Bildung und Erziehung in den USA. Insgesamt sind jedoch die Anpassungen, die in konkreten Anwendungsfällen vorgenommen wurden, eher geringfügiger Art. Die Grundstruktur der ‚Standards‘ mit den vier Hauptgruppen ‚Nützlichkeit‘, ‚Durchführbarkeit‘, ‚Korrektheit‘ und ‚Genauigkeit‘ ist in keinem der uns bekannten Fälle verändert worden. Auch die überwiegende Mehrheit der einzelnen ‚Standards‘ ist in allen Anwendungen beibehalten worden. Die vorgenommenen Änderungen betreffen jeweils einige wenige Einzelstandards, die geändert, mit anderen Standards zusammengeführt oder im Einzelfall ganz gestrichen wurden.

Zusammenfassend empfehlen wir aufgrund der bisherigen Erfahrungen die Anwendung der ‚Standards‘ auf Evaluationen in Deutschland, Österreich und der Schweiz, und zwar auch dann, wenn Evaluationen außerhalb des Bereiches Bildung und Erziehung durchgeführt werden.